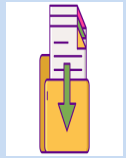
Construction Re-engineering Presentation to CSAC

March 19, 2021

Build BPS with 3rd Party Data Sources

Collection

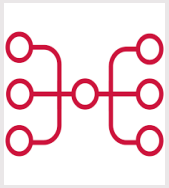


Data files from vendors

Identify cut-off dates and select files for October and November

Processing

Filter data for
1. New SFH
2. Issued Status



Jurisdictions mapping from Vendors to Census BPS
1. Manual mapping by name
2. Shape files mapping

Categorize construction type and aggregate for one unit
- Cabins
- Duplexes & Twin Homes
- Single Family Homes

Compare one unit data between BPS and vendors

Analysis

Selected around 100 Jurisdictions which have difference greater than 5% or 50 units between BPS and vendor data

Create profiles by Jurisdictions to identify best matches to BPS

Dissemination

Share the findings with Census team

Continuous collaboration with the Census Survey Stats & Math Stats Team

Data Collection

Construction Monitor

Weekly data delivery on every Monday through FTP starting Oct 13th → Select cut-off dates and Pick files from week 41-47 for October and week 45-51 for November

CM	Jurisdictions covered	Records count
Oct	1643	164 k
Nov	1485	128 k

CoreLogic

Weekly data delivery on every Thursday through FTP starting Nov 16th → Select cut-off dates and Pick files for October and November

CL	Jurisdictions covered	Records count
Oct	458	137 k
Nov	201	88.5 k

The jurisdictions covered by 3rd party data sources account for about 70% of single-family units authorized in the U.S.

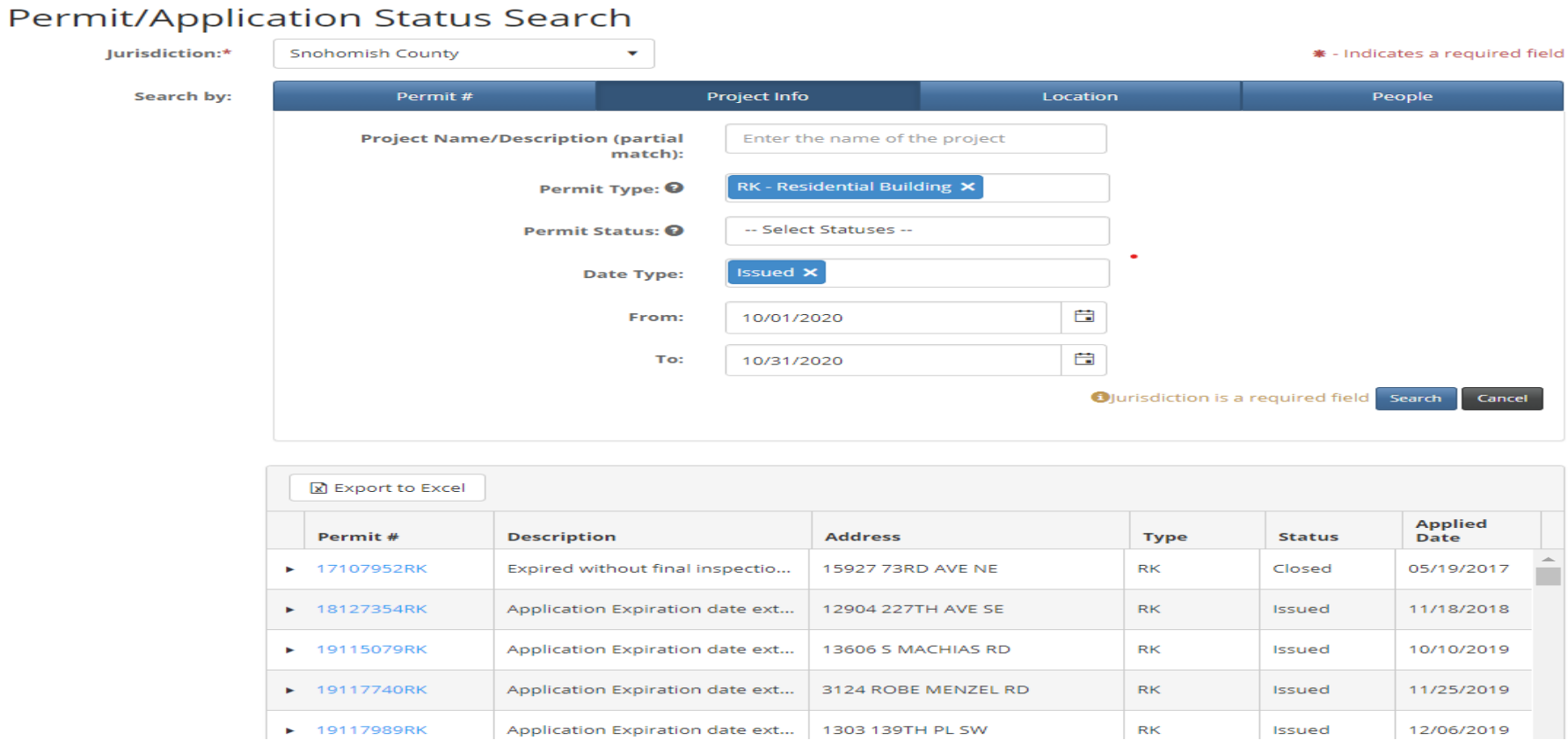
Data Collection

Website

Search if BPS data is available online

→

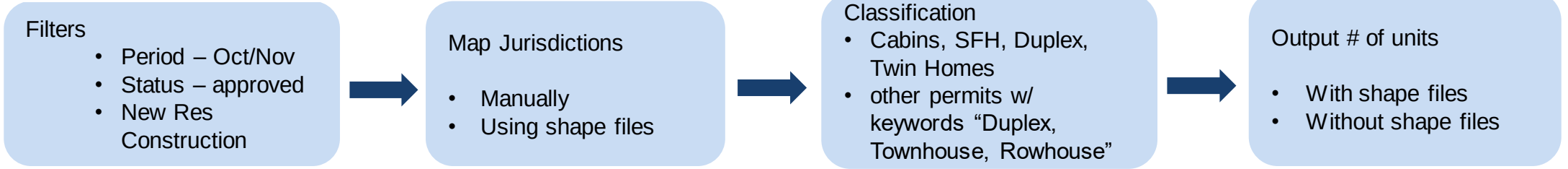
Export CSV/ excel if available, otherwise scrape data (API/PDF) for Oct/Nov



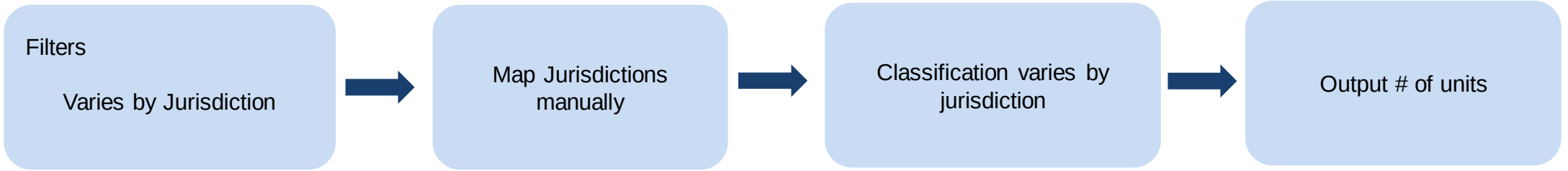
Permit #	Description	Address	Type	Status	Applied Date
17107952RK	Expired without final inspectio...	15927 73RD AVE NE	RK	Closed	05/19/2017
18127354RK	Application Expiration date ext...	12904 227TH AVE SE	RK	Issued	11/18/2018
19115079RK	Application Expiration date ext...	13606 S MACHIAS RD	RK	Issued	10/10/2019
19117740RK	Application Expiration date ext...	3124 ROBE MENZEL RD	RK	Issued	11/25/2019
19117989RK	Application Expiration date ext...	1303 139TH PL SW	RK	Issued	12/06/2019

Data Processing

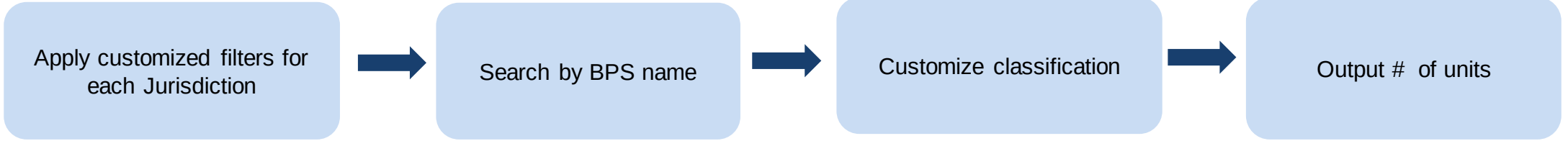
Construction Monitor



CoreLogic



Website



Next Steps (….over next 3-4 months)

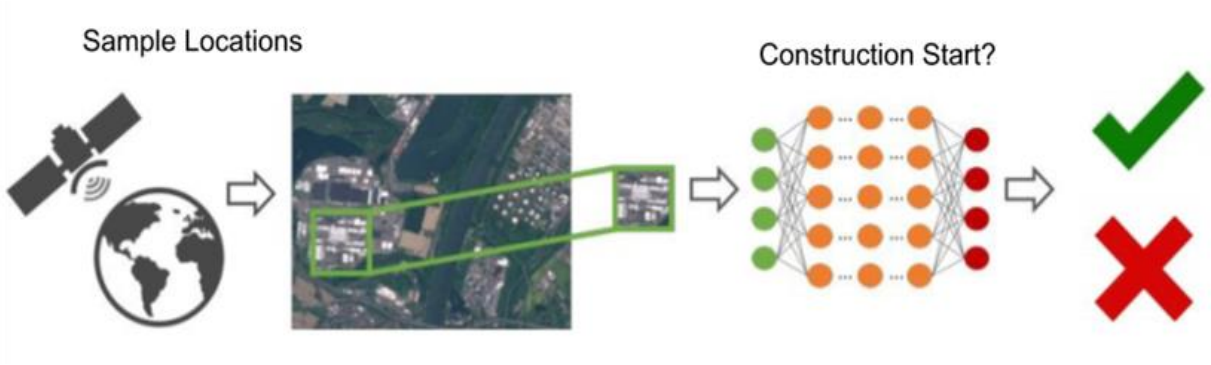
- Expansion of jurisdictions
- Jurisdiction mapping based on Census geolocations
- Classification of construction types library
- Add Multi-family homes to BPS work stream

Applying and AI and Satellite Imagery for Construction Classification

Objective and Success Criteria

Goal

- Demonstrate potential value of the identification and categorization of construction sites using satellite images.



Definition of construction starts:

- *Excavation and delimitation of building site*
- *Laying the concrete foundation of the building*

Success Measures

- Cost: Demonstrate that construction starts and completions can be accurately estimated using satellite imagery at a cost equal to or less than existing methods
- Timeliness: Demonstrate through the POC that monthly building starts can be estimated using geospatial images and ML/AI as per existing program timelines

Initial Project Phases

November 2019 to April 2020

Research and Experimentation

- Collaborate with StatsCanada
- Research ML models
- Experiment on free construction image dataset
- Define POC requirements

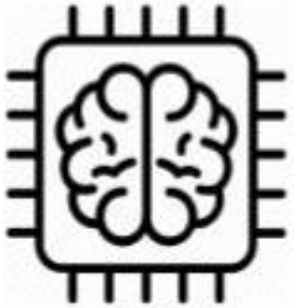
May to July 2020

Image Search and Collection

- Filter permit data
- Select vendor
- Image search
- Image ordering and downloading

August to September 2020

Model Development and Training



- Develop Deep Learning Models
- Train, test, and improve

September > 2020

Validating and Future Improvements

- Validating results
- Monthly production
- Training of new models for advanced classifications

Image Categorization

Pre-constructions
90 days before permit authorization date or earlier



This image category displays the construction stage before any construction has happened. Images display the ground untouched and no major delimitations or excavations.

Construction Starts
Between 30 days after permit authorization date and 125 days after it.



This image category displays the construction stage immediately after construction has started. Our defined “start” definition is: “visible excavation or foundation”.

Construction Completions
270 days after permit authorization date.



This image category displays the completed roof covering the area where there was a foundation or excavation previously.

CNNs Development Mode

Model A (FastAI)

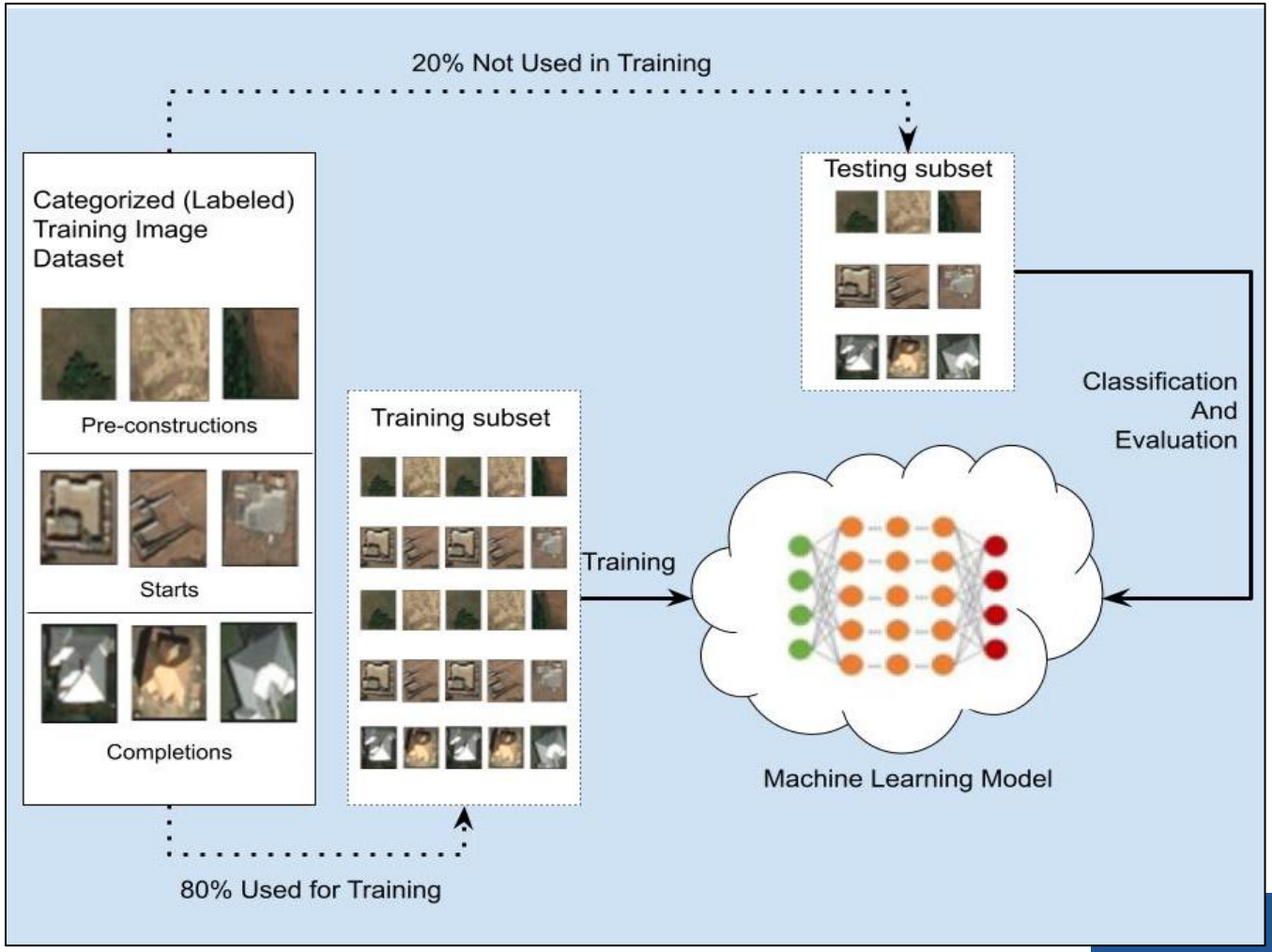
We utilized Fast.ai libraries and ResNET, a pre-trained Deep Learning model for image classification most applied to analyzing visual imagery. StatsCanada has shown promising results using similar methodology and tools.

- Rapid Prototyping
- Best utilization and optimization for GPUs

Model B (Keras)

Keras is a deep learning API written in Python, running on top of the machine learning platform TensorFlow. It was developed with a focus on enabling fast experimentation. Being able to go from idea to result as fast as possible is key to doing good research.

- Large-scale deployments
- Cross-platform



Training and Testing

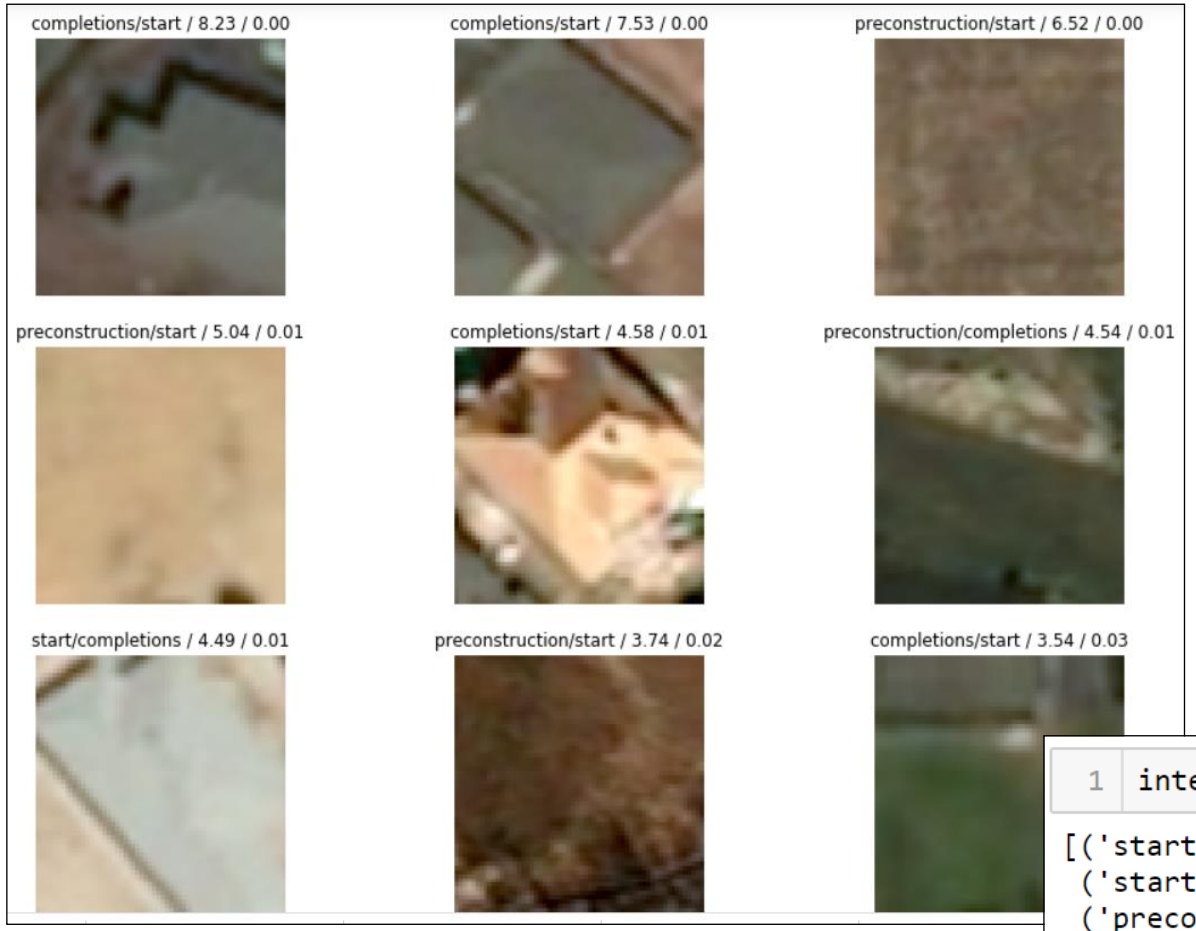
The **FastAI model** was trained to classify between 3 construciton stages image categories (precons, starts, and completions). It has been trained with 3236 images in the training subset and 808 images in the relative validation subset. Initial values of the fitting, showed 8% error rate (or 92% accuracy score). ***After optimizing the FastAI model we achieved 94% accuracy score***.

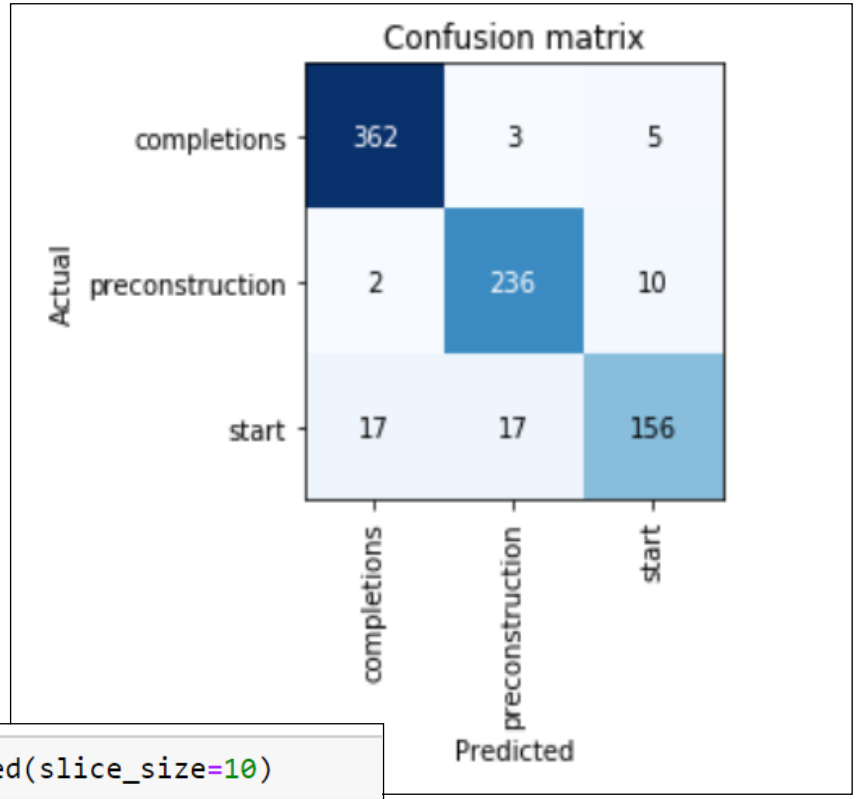
epoch	train_loss	valid_loss	error_rate	time
0	0.114260	0.310744	0.084158	00:26
1	0.172425	0.280688	0.081683	00:27
2	0.158284	0.250213	0.077970	00:29
3	0.123952	0.245381	0.082921	00:29

For **Keras model**, the team used 3,262 images as the training set and 814 images as the validation set. The images are labeled as “Complete” and “Incomplete”. ***The model has achieved an accuracy score as high as 93%*** during the training process, with an average accuracy score around 90%. As we further train/ tune the model with more images from our vendors, we will be able to continue to improve our model performance.

epoch	train_loss	valid_loss	error_rate	time
0	0.100892	0.211188	0.077970	00:24
1	0.099556	0.206739	0.075495	00:25
2	0.096449	0.208337	0.077970	00:24
3	0.097957	0.205757	0.066832	00:24

Validating Models






```
1  interp.most_confused(slice_size=10)
```

```
[('start', 'completions', 17),
 ('start', 'preconstruction', 17),
 ('preconstruction', 'start', 10),
 ('completions', 'start', 5),
 ('completions', 'preconstruction', 3),
 ('preconstruction', 'completions', 2)]
```

Current Project Phases

November 2020 to December 2020

Location Analysis and Input Data

- Preprocess NP and SOC Starts data and add GeoAnalysis
- Validating shapefiles to each place
- Aggregating construction start records for image search

January 2021

Image Search and Collection

- Filter by starts density
- Search parameters
- Collection methodology
- Gridding original image into tiles of smaller area

February 2021

Production Mode and Classifications

- FastAI v2 upgrade
- Model retraining
- Hunting Mode Implementation
- Classifying multiple images

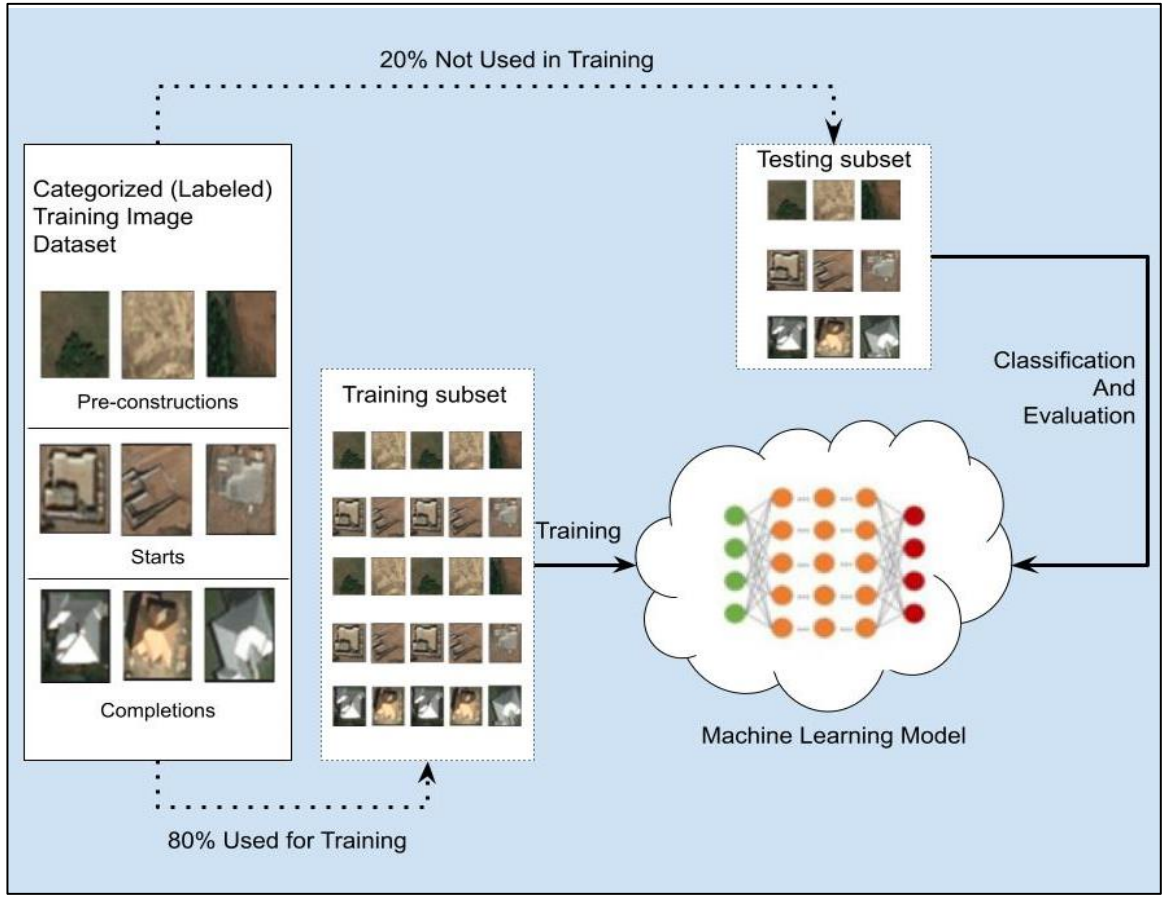
March > 2021

Validating and More Specific Training

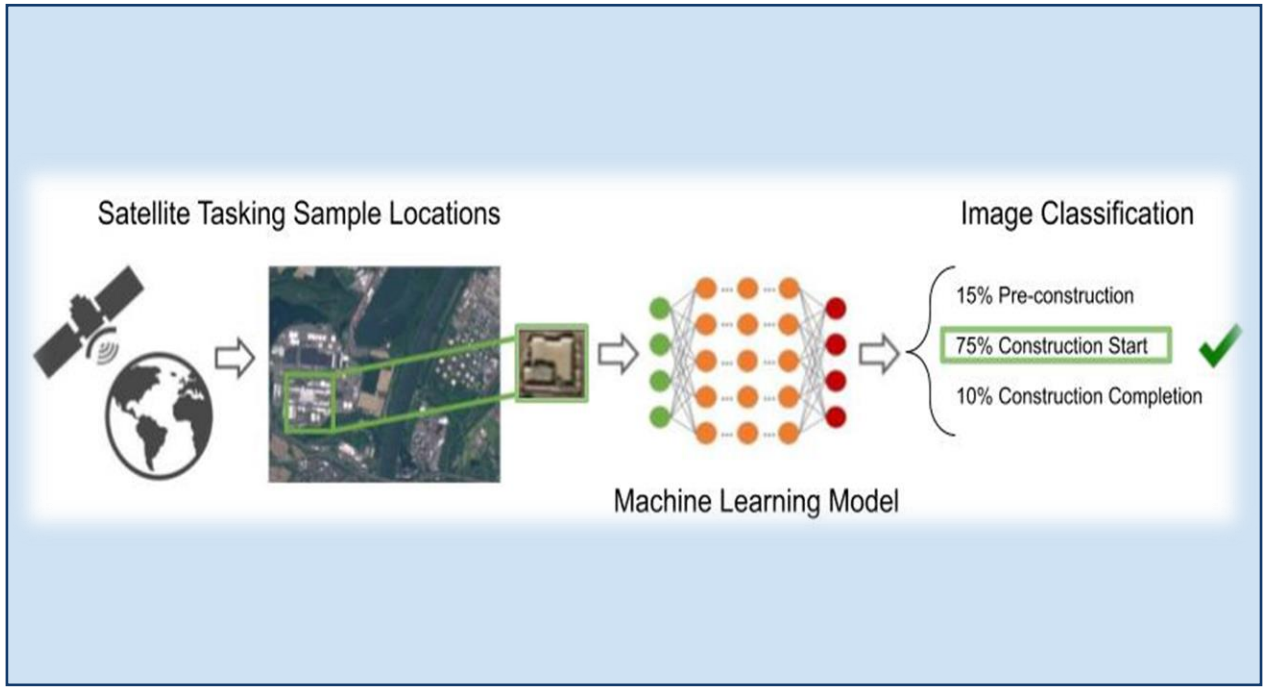
- Evaluating Classification Metrics
- Collab with StatsCanada and CSAC
- Acquire more images and increase validation scale

From Dev to Production

Training and Testing Model



Model used in Hunting Mode

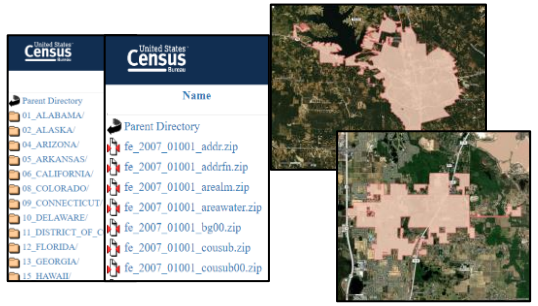


Mapping Construction Starts Data

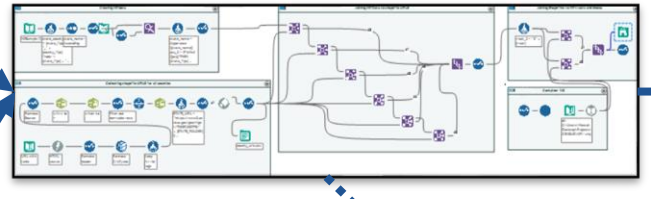
- Pre-process Construction Activity Data
- Shapefile Matching and Spatial analytics

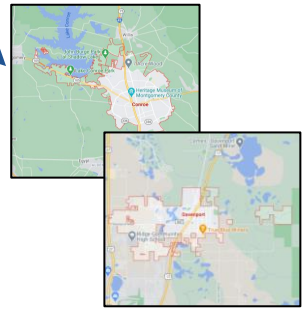
SOC Data
(Starts count per location)

Census GEO Shapefiles



Automated Alteryx Workflows





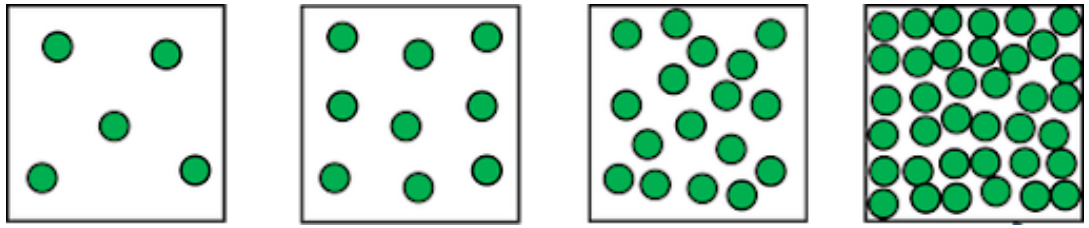
- Validating w/ Google Maps

- Shapefiles matches
- Aggregating records by location, data, and construction starts count

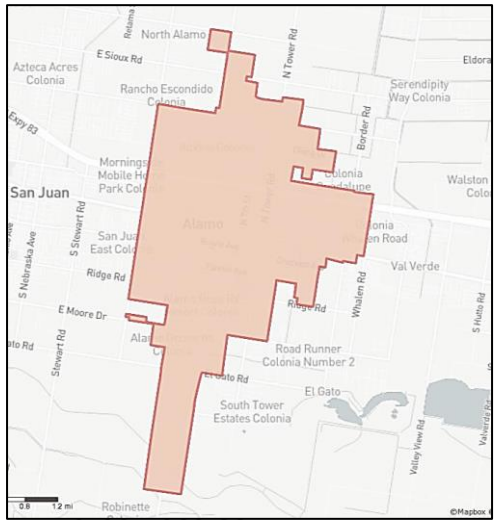
SOC Starts
Locations Mapping

Image Data Collection

- Filters by place construction density: (starts count / area in sq km)



- Bounding Boxes



- Image search parameters

< 15% Cloud coverage

Image fully covers shapefile

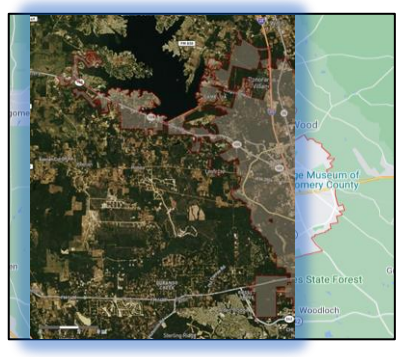
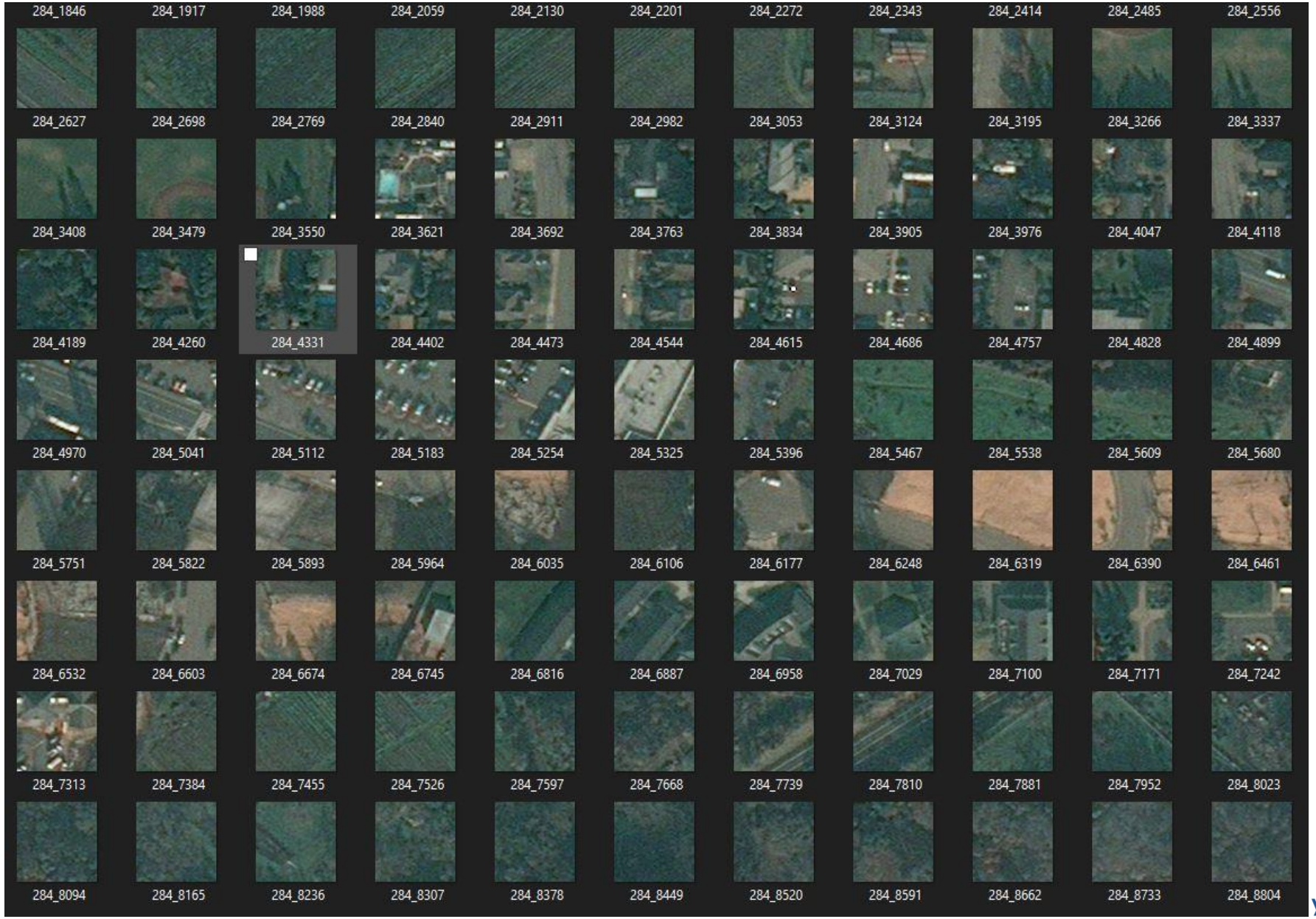


Image Collection & Grid Methodology

- Image search selected sample

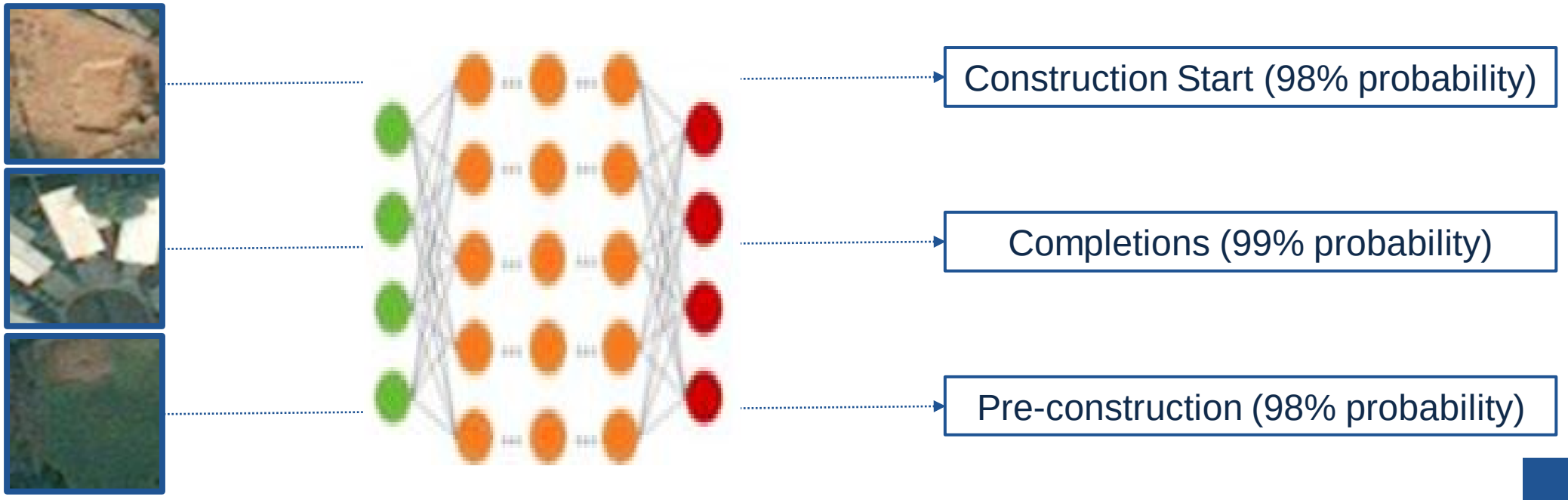
- Gridding into 0.001 sq km tiles

Avg construction lot size = 8900 sq ft, + margin = = 0.001 sq km



Production Mode

- FastAI v2 on a new **dedicated** Reveal Linux server
- Model re-trained on server (GPU enabled)
- Creating production mode environment: from grids to answer

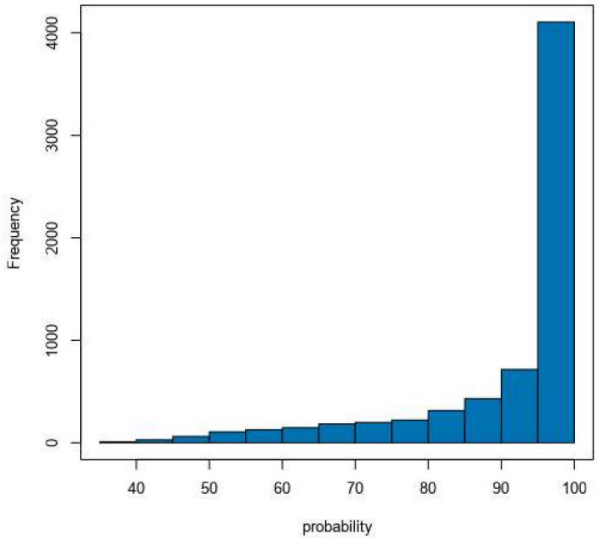


Construction Classifications

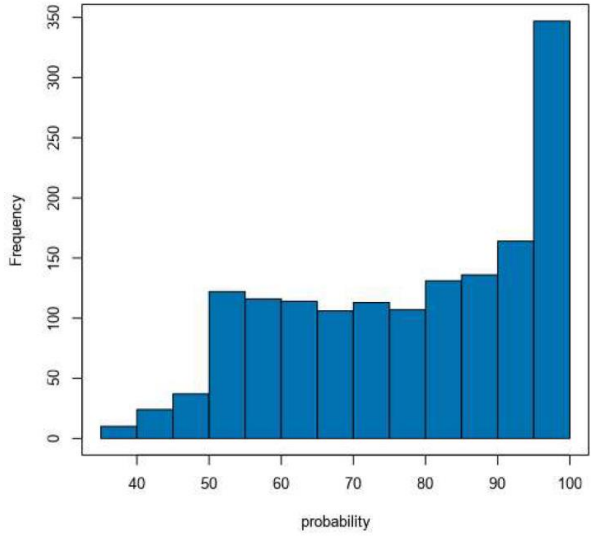
- Classification Process and Metrics
  - 9409 tiles (images) classified
  - 13 minutes classification time

Category	Avg_probability	Max_probability	Min_probability
Completions	82.526856	99.9999	36.6909
Preconstructions	91.195509	99.9998	35.5002
Start	77.87969	99.9999	35.6627

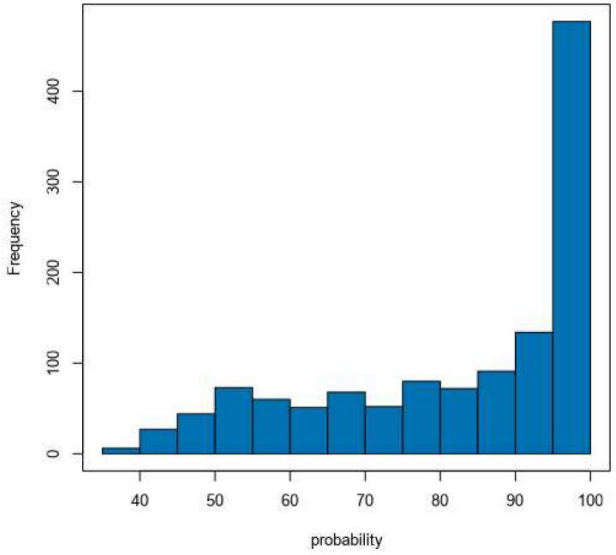
Preconstructions



Starts



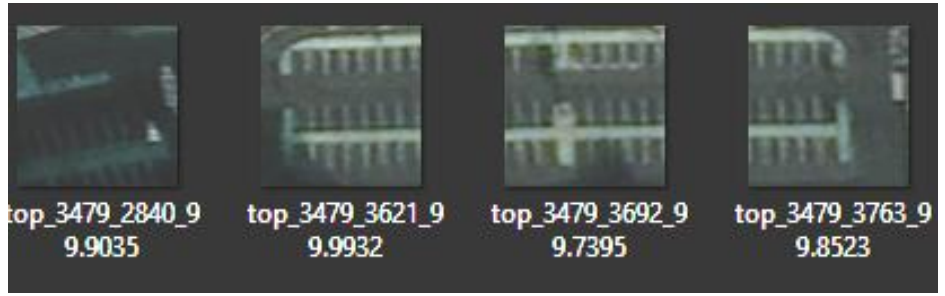
Completions



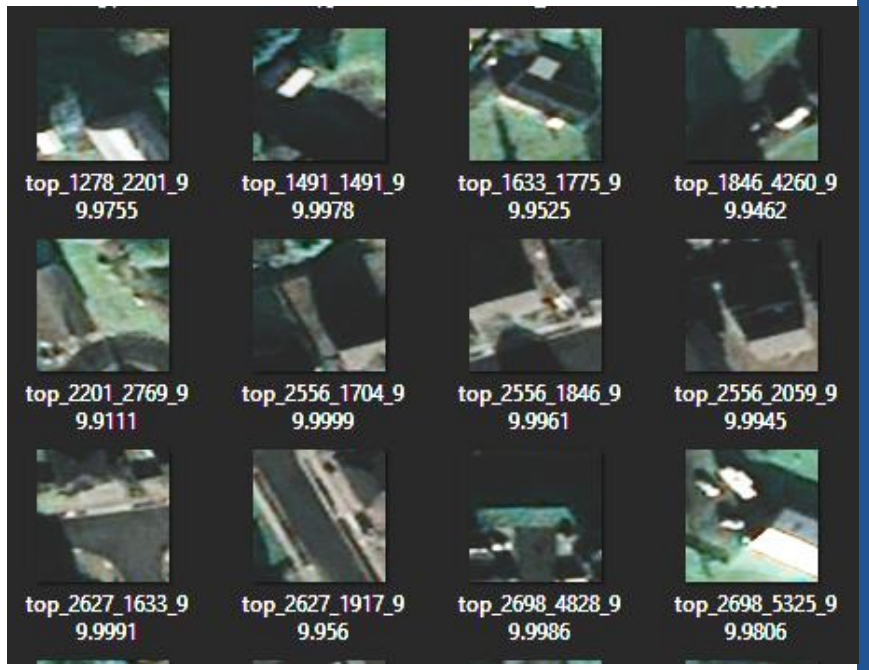
United States Census 2020

Precons

Starts



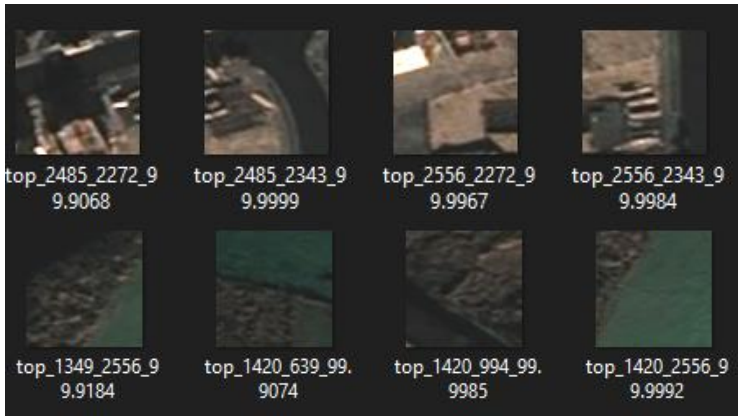
Completions



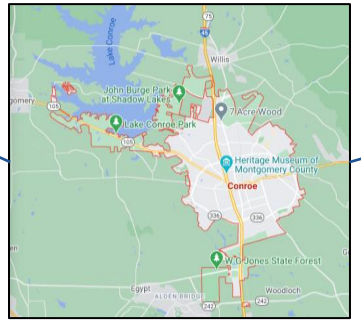
Hunting Mode Validation

- Taking results from classifications
- Aggregating based on locations (geoid)
- Matching aggregation against Census Data

X Image Classification Starts Count ⟷ *(X / Y) % match?* ⟷ *Y Census Starts Count*



Sample Location

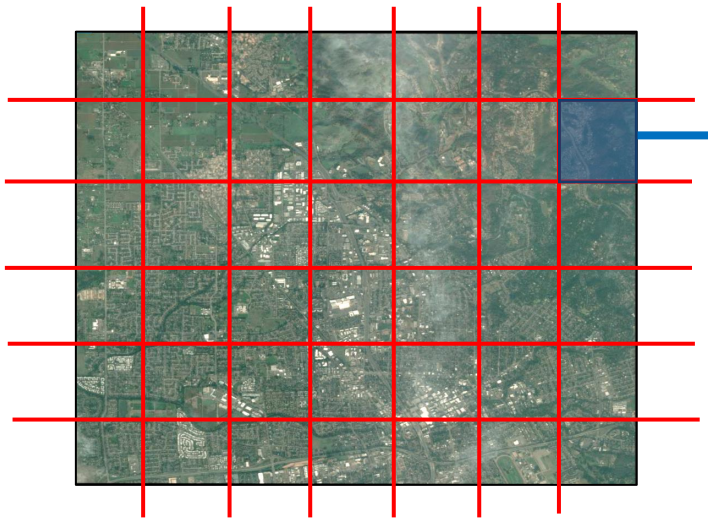


SOC Data
(Starts count per location)

Revised Grid Methodology

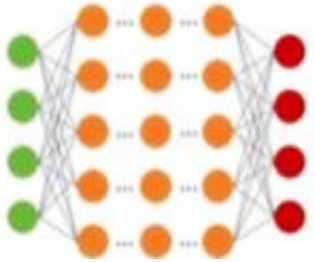
- Collect image for selected location

- Grid to bigger tile area of 0.1sqkm



- Automatic Construction Boundary Detection

Classification Model

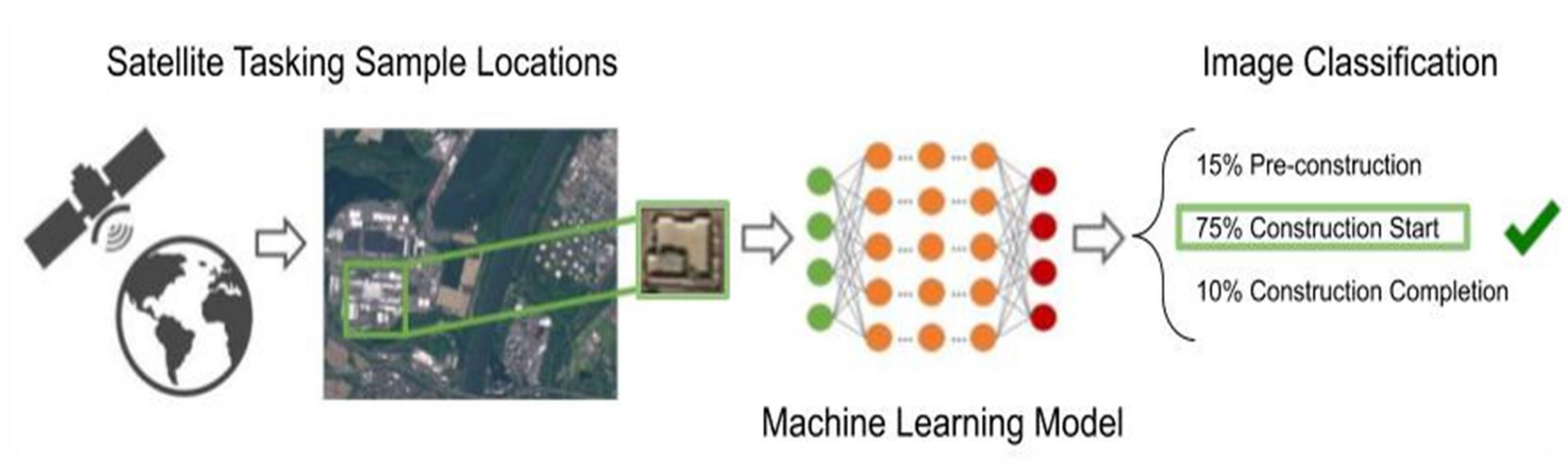


Start

Completion

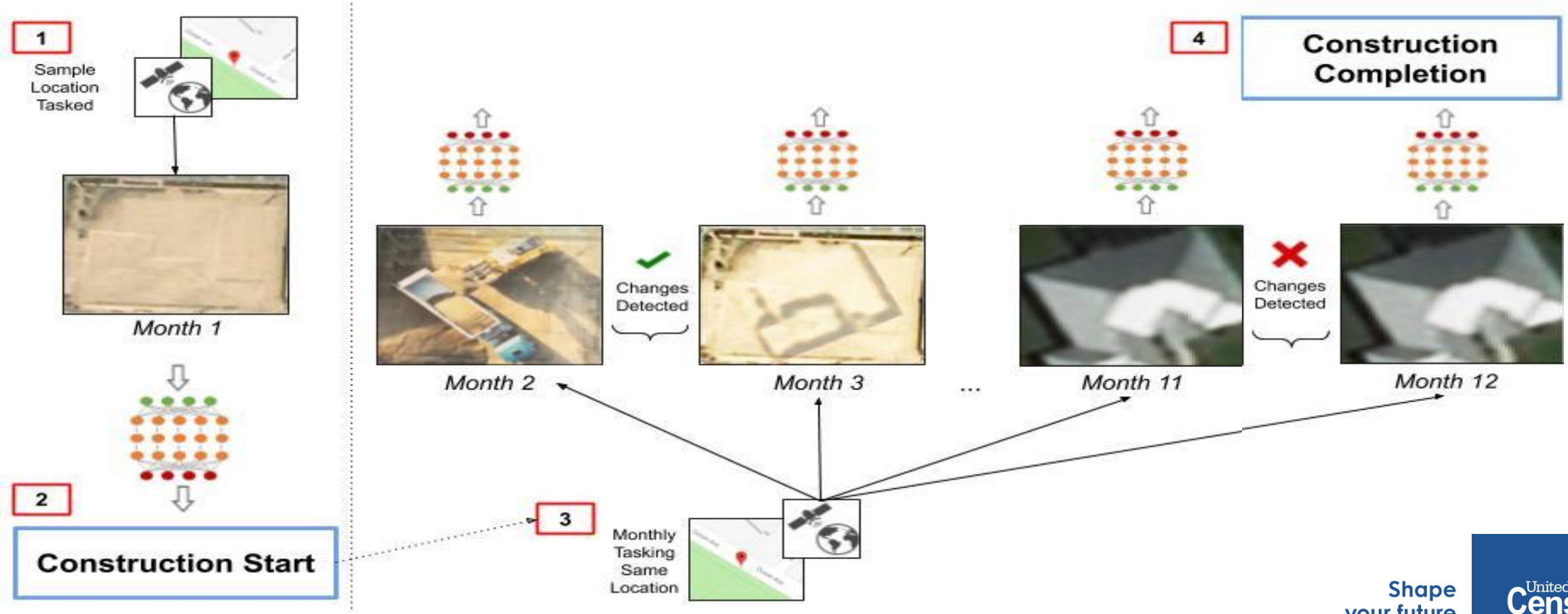
Hunting Mode

After refining our training image dataset, we decided to use two modalities for operations of the satellites. In “Hunting Mode”, images are collected monthly and fed to the CNN models to detect if there are any construction starts.



Tracking Mode

Once a construction start has been detected in an image, we will track that location on a monthly basis for changes. We will apply change detection techniques to track construction progress and tag the construction activity as complete based on the detection of pre-established features such as finished roof, driveway, landscaping, absence of construction vehicles, etc.



Change Detection Implementation

Change Detection algorithm in between construction stages is already implemented and tested, ready for tasked images.

Starts

Completions

Next Steps

- GUI interface for human validation and model feedback
- Classify more SOC Starts locations and test Non-Permit locations
- Automatic construction boundary detection
- Training new non-construction categories to filter out false positive starts:
  - Streets, parking lots, trees, pools
- Specialize current construction categories:
  - urban starts, rural starts
- Integrate Hunting and Tracking mode
- Train model on residential multi-unit, and on commercial buildings
- Train model to infer construction $ value